Short Paper: A video self-avatar influences the perception of heights in an augmented reality Oculus Rift

M. Gutekunst[1] & M. Geuss[2] & G. Rauhöft[1,2] & J. Stefanucci[3] & U. Kloos[1] & B. Mohler[2]

[1]Reutlingen University, Reutlingen, Germany
[2]Max Planck Institute for Biological Cybernetics, Tübingen, Germany
[3]University of Utah, Salt Lake City, USA

Abstract

This paper compares the influence a video self-avatar and a lack of a visual representation of a body have on height estimation when standing at a virtual visual cliff. A height estimation experiment was conducted using a custom augmented reality Oculus Rift hardware and software prototype also described in this paper. The results show a consistency with previous research demonstrating that the presence of a visual body influences height estimates, just as it has been shown to influence distance estimates and affordance estimates.

Categories and Subject Descriptors (according to ACM CCS): H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems —Artificial, augmented, and virtual realities

1. Introduction and previous work

Humans are most familiar with the size and dimensions of their own body, leading to interesting units of measurements such as ells or feet. As such it can serve as a valuable point of reference when estimating distances or heights and improves the precision with which tasks related to the participant are executed, i.e. distance estimates [RIKA08, MCRTB10].

To view heights while keeping subjects safe, Gibson and Walk [GW60] introduced the visual cliff paradigm, where the subject is always on solid ground but can see a hole below them. Originally this was achieved with a glass floor, yet can be transferred to virtual reality by using a head mounted display (HMD). Meehan et al. [MIWB02] first used the visual cliff paradigm in virtual reality to investigate if physiological responses (heart rate, perspiration and body temperature) to a stressful environment would be similar to what is known to occur at the edge of an actual cliff. Usoh et al. [UAW*03] extended this work by adding a virtual character-based body to the experience and investigating navigation and presence in the visual cliff paradigm. Ries et al. [RIIK10] have also explored the benefits that self-embodied virtual avatars provide to a user's sense of presence in the visual cliff paradigm. Lin et al. [LRB13] demonstrated that having a virtual avatar influenced height estimates and expectation of action performance with regard to stepping off a ledge.

Previous research has investigated the perception of heights in both real [SP12] and virtual environments [Geu14]. Participants estimated heights by performing a perceptual matching task whereby participants adjusted a reference distance to match the vertical extent. Stefanucci and Proffitt (2009) have shown that heights are overestimated by 60% when viewed from above and 30% when viewed from below. In VEs, heights are similarly overestimated when viewed from above [Geu14]. Specifically, Geuss [Geu14] found that participants, on average, overestimated virtual heights by 75%. The greater degree of overestimation in the VE when compared to the real environment may be due, in part, to a lack of a visual body in the VE.

VEs are useful because variables of the environment can be changed with minimal effort, such as depths of pits or distances of objects. However considerable effort is required to integrate a virtual body into the VE. Often the integration of a virtual body is achieved by tracking the person's movements and mapping them to an avatar [LNWB03, MCRTB10, RCRR*12, RIKA08, RIIK10, UAW*03]. The issue with having a character based self avatar in VE is that the body one sees is only a virtual representation of the real body and will likely not be quite as familiar. Another approach is to have a video see through self-avatar using cameras mounted to the head-mounted display [FI12].

We describe the implementation of a video-based self avatar, using concepts of augmented reality to project one's body into the VE, allowing one to see her own body from one's point of view, as would be natural in the real world. One benefits of this approach is the simplicity to integrate a realistic self-avatar into a VE by extracting visual data from the video feed provided by the cameras and not requiring expensive tracking equipment. This approach also saves time by offering a direct feed of the real body of the user and does not require a body scan to provide a realistic representation [PSR*14]. The prototype can be placed within augmented virtuality according to the reality-virtuality continuum presented by Milgram et al. [MTUK95].

The main questions of our study are whether the use of video based self avatars alters one's perception and estimation of heights. We want to test how a video based self avatar can help to improve the estimation of heights. The implementation of the approach was hereby created as a simple hardware and software prototype using only publicly available applications and appliances. For the evaluation we have performed a height estimation experiment described in this paper.

We leave out of the scope of our project object interaction (such as described by Lok et al. [LNWB03] and Raj et al. [RCRR*12]), as the only object for this paradigm we imagine interacting with are objects that the user would occlude (i.e. a virtual railing or the edge of the cliff).

2. Technical description

For the purpose of this research we created a camera based HMD prototype based on the Oculus Rift Dev. Kit 1 (OR) that was designed to allow the augmentation of a virtual environment by integrating real world content. The basic idea for this prototype is to make it possible for a person to see their body within a virtual environment without the need of a programmer to create a character based self-avatar, animated based on the tracked movement of the person. The benefit of this approach is that a person will be able to see their own body, including clothing and jewelry, with the exact movement and shape of the own body. In this way we hope to provide a more realistic view of one's body and to increase the amount of self-awareness. The prototype captures the real environment surrounding a person, then extracts the relevant parts of the image (in our scenario: the hands and arms of the person) and displays these parts within the virtual environment. To do so, the prototype consists of a stereo camera set-up, the OR and an in-house developed application to perform the image processing.

2.1. Hardware Specification

The hardware setup contains the OR as well as two camera mounts with custom built cameras. The cameras and the mounts were built on a set-up described in detail by Steptoe [Ste13]. The cameras consist of a Logitech C310 Webcam and the lens of a Genius Widecam F100. They support a maximum resolution of 1280x960 pixels, have a 150 degree FOV and an aspect ratio of 1.33:1. Steptoe estimated the maximum frame rate of the C310 to be about 50Hz [Ste13]. The cameras are mounted to the OR using spherical joints

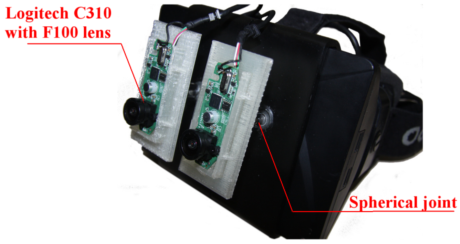

Figure 1: *Oculus Rift with the mounted cameras*

and are therefore highly adjustable. Each camera mount additionally holds a little head-light (a 3V warm-white LED below the camera) which is not visible in the image above. Since the OR does not provide absolute information about head position and rotation we instead used a motion capture system with 16 Vicon MX13 cameras. This was done to enable the participants to look down into a pit.

2.2. Software

The application was developed as a plugin for the Unity3d engine. The plugin is called every 10th frame in Unity. It captures the current frame from each of the cameras and processes it. The processed image from each camera is then applied to a texture in Unity which is used to combine the contained video data with the visual data from the virtual world.

First the image is segmented to extract the relevant information. Several segmentation approaches were considered. A feature based segmentation where the features in the image were used to classify the objects within the image was considered, but the hands of a person did not provide enough features for a robust classification. Using a histogram based color model turned out to be too slow when using a sufficient amount of histograms. We decided to use a color model built up manually by an expert defining a range of colors to be either included or excluded in the processed image. This procedure was performed on every testing day before the experiment. This approach turned out to be the fastest and most robust of the evaluated alternatives. Using the color model an alpha mask of the image is created. This mask is used to extract the relevant data from the image. After removing noise and small objects, the remaining objects are classified. The primary classification criterion is the existence and amount of convexity defects, which can be an indicator for fingers and therefore, a hand. The convexity defects are estimated using the convex hull of an object. The second criterion is defined by the amount of points that form the convex hull. Only the objects that match all criteria are extracted from the image by using the alpha mask. The objects are then resized and repositioned in relation to the output resolution and their original attributes. Finally the extracted objects are applied to the Unity-textures using OpenGL.

2.3. Issues and limitations

A major issue in the current set-up is the frame rate. The image segmentation causes the frame rate of the virtual world to drop from about 710 FPS without the plugin, to about 136 FPS. Of course this is still more than the OR can handle but the frame rate of the video feed that is displayed in the virtual environment is much lower (about 14 FPS), because the plugin is called after each 10th rendered frame. Another issue is that the color based image segmentation is not robust with complex backgrounds. When other objects with related colors and shapes are in the scene, they might be recognized as the body and displayed in the virtual environment. To counteract this behavior the previously mentioned head-lights were introduced to improve the contrast between foreground and background. In the experimental set-up we also darkened the room to increase this effect. Another limitation in the current technical setup is that there is no way to obtain the 3D position of the persons body-parts from the real world. It was attempted to create a disparity map for depth estimation, but all tested methods are either not precise enough or too slow for a real time application. Due to this we are missing 3D coordinates in the virtual world and therefore interaction with virtual objects is left for further work.

3. Experiment

A within-subject design experiment with 16 participants (8 male, 8 female; mean age 29.1 years) was conducted to investigate the influence of a video self-avatar on the estimation of heights at a visual cliff in an augmented OR HMD setup. We varied the depth of the pit as well as the visibility of the body (counterbalanced for order).

3.1. Experimental set-up

In the experimental set-up a scenario was used in which participants stood on a virtual plank over a pit (an enclosed visual cliff with dimensions 4 x 4 meters) and were asked to estimate the depth of the pit below them. We measured the accuracy of the estimated height for each task. The experiment was split in to two counterbalanced conditions: The first condition restricted any visual perception of the surrounding reality. The second condition had the prototype's cameras switched on, to enable the participant to perceive her or his own body. The experiment was implemented using Unity 4.3 and was run on a computer using Windows 7 Enterprise, 64 bit. The system had an Intel Xeon E5-1620 Processor running at 3.6 GHz, 16GB of memory and an NVIDIA GeForce GTX 760.

3.2. Experimental flow

The participants were reminded to use their hands for reference purposes if it felt natural to them. The use of a controller as an input device further helped participants to keep their hands in front of them and therefore in view. Every participant viewed their hands at least once at the beginning of the condition for at least 5 seconds. Each condition was separated into two parts: During the introduction stage, the pit was covered up and the participants were requested to estimate the **distance** between a reference ball (with a shadow on the ground surface below) with a radius of 10 cm, hovering 40 cm above the virtual ground to a marker lying just at the edge of the pit in front of the participant. Input occurred by moving a ball located to the left or right (in an alternating fashion) of the participant to the perceived distance of the reference ball using the controller. Each participant repeated the task six times with distances incrementing from 5 meters to 10 meters inclusively, in steps of 1 meter. The second part of the experiment occurred with an open pit (see Figure 2). The participants adjusted the distance of the ball according to the depth of the pit below them. The adjustable ball and marker were the same as in the first part of the experiment, except that they *randomly* appeared either to the right or the left of the participant. The depth of the pit and order of the depths were kept the same for every subsequent participant, starting at a depth of 1 meter and ranging to 15 meters in steps of 2 meters. The range of the depths was repeated three times for every participant and every condition, resulting in 24 depth estimations per condition.

Figure 2: *The condition where the participant is able to see the own body in virtual environment.*

4. Results

The mean relative errors were calculated as the estimated heights subtracted by the actual heights divided by the actual heights, are displayed in Figure 3. A 2 Condition (visible, invisible) x 2 Order x 8 Height x 3 Repetitions repeated-measures ANOVA was run on the mean relative errors. The only between-subjects factor was the order. The within-subjects factors were condition, the height and the repetition. Analysis revealed an interaction between Condition and Order, $F(1,14)=12.762, p=.003$. For the Order invisible then visible participants overestimated heights more when the hands were invisible (M=139.90%, SE=19.8%) than when the hands were visible (M=111.03%, SE=17.7%), $F(1,7)=14.389, p=.007$. For the Order visible then invisible estimates did not differ between invisible (M=77.30%, SE=11.5%) and visible (M=84.90%, SE=12.5%), $p>.05$. This results suggest that seeing one's arms improves estimates of heights.

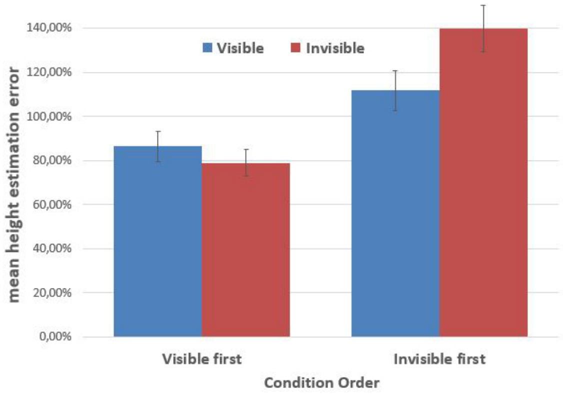

Figure 3: *Mean error on height estimation separated by groups and conditions with arms visible and invisible for the different group orders. Error bars represent one standard error of the mean.*

5. Discussion

Using an augmented reality OR which enabled people to see their own body while viewing a virtual cliff, we found that people estimated the heights of the cliff more accurately when their body was visible. This finding demonstrates the importance of the visible body for perception. These results are consistent with research that demonstrated a virtual body improved the accuracy of distance estimates in VEs [RIKA08, MCRTB10] but extends this work to heights.

We also found that the order of presentation of the body (visible or invisible) had an effect on the height estimates. This could be because people wanted to be consistent with their height estimates over time or the ability to see one's body may have a lasting effect on subsequent height judgments.

These results are promising for applications where virtual reality can treat acrophobia [HKM*95, CWHW09], suggesting that seeing one's body gives people a more accurate perception of the height at the cliff. This might also enable scientists to further explore the root causes and triggers for acrophobia that relate to the visually seen body i.e. postural stability, self-motion and visual control of one's body.

This research raises several important questions for future research. Did people that reported higher levels of presence in the space more accurately estimate heights and distance. Did we get consistent results with previous scientists who found that physiological measures were elevated at the edge of the cliff and more so when a video self-avatar was seen [MIWB02, UAW*03, RIIK10]? Future research could focus on whether the visual body needs to be a self-avatar, or is it only important to have a familiar size reference of a human body to help scale the space. Fortunately, since virtual and augmented reality technology makes it rather trivial to manipulate elements of the virtual world we can answer all of these questions which would be challenging, time-consuming or impossible without the use of this technology.

References

[CWHW09] Coelho C. M., Waters A. M., Hine T. J., Wallis G.: The use of virtual reality in acrophobia research and treatment. *Journal of Anxiety Disorders 23*, 5 (2009), 563 – 574. 4

[FI12] Fiore L. P., Interrante V.: Towards achieving robust video self-avatars under flexible environment conditions. *International Journal of Virtual Reality (special issue featuring papers from the IEEE VR Workshop on Off-The-Shelf Virtual Reality) 11*, 3 (Sept. 2012), 33–41. 1

[Geu14] Geuss M.: *An investigation into dynamic changes in visually-guided actions and perceived spatial layout when afraid.* PhD thesis, University of Utah, August 2014. 1

[GW60] Gibson E. J., Walk R. D.: *The" visual cliff".* WH Freeman Company, 1960. 1

[HKM*95] Hodges L., Kooper R., Meyer T., Rothbaum B. O., Opdyke D., de Graaff J., Williford J., North M.: Virtual environments for treating the fear of heights. *Computer 28*, 7 (Jul 1995), 27–34. 4

[LNWB03] Lok B., Naik S., Whitton M., Brooks F. P.: Effects of handling real objects and self-avatar fidelity on cognitive task performance and sense of presence in virtual environments. *Presence: Teleoperators and Virtual Environments 12*, 6 (2003), 615–628. 1, 2

[LRB13] Lin Q., Rieser J. J., Bodenheimer B.: Stepping off a ledge in an hmd-based immersive virtual environment. In *Proceedings of the ACM Symposium on Applied Perception* (New York, NY, USA, 2013), SAP '13, ACM, pp. 107–110. 1

[MCRTB10] Mohler B. J., Creem-Regehr S. H., Thompson W. B., Bülthoff H. H.: The effect of viewing a self-avatar on distance judgments in an HMD-based virtual environment. *Presence: Teleoperators and Virtual Environments 19*, 3 (2010), 230–242. 1, 4

[MIWB02] Meehan M., Insko B., Whitton M., Brooks Jr. F. P.: Physiological measures of presence in stressful virtual environments. In *Proc. ACM SIGGRAPH* (2002), pp. 645–652. 1, 4

[MTUK95] Milgram P., Takemura H., Utsumi A., Kishino F.: Augmented reality: A class of displays on the reality-virtuality continuum. In *Photonics for Industrial Applications* (1995), International Society for Optics and Photonics, pp. 282–292. 2

[PSR*14] Piryankova I. V., Stefanucci J. K., Romero J., De La Rosa S., Black M. J., Mohler B. J.: Can i recognize my body's weight? the influence of shape and texture on the perception of self. *ACM Trans. Appl. Percept. 11*, 3 (Sept. 2014), 13:1–13:18. 2

[RCRR*12] Raj M., Creem-Regehr S. H., Rand K. M., Stefanucci J. K., Thompson W. B.: Kinect based 3d object manipulation on a desktop display. In *Proceedings of the ACM Symposium on Applied Perception* (2012), ACM, pp. 99–102. 1, 2

[RIIK10] Ries B., Interrante V., Ichniowski C., Kaeding M.: Investigating the physiological effects of self-embodiment in stressful virtual environments. *IEEE VR 2010 Workshop on Perceptual Illusions in Virtual Environments)* (2010)), 176–198. 1, 4

[RIKA08] Ries B., Interrante V., Kaeding M., Anderson L.: The effect of self-embodiment on distance perception in immersive virtual environments. *ACM Symposium on Virtual Reality Software and Technology* (2008), 167–170. 1, 4

[SP12] Stefanucci J. K., Proffitt D. R.: The roles of altitude and fear in the perception of height. *Journal of Experimental Psychology: Human Perception and Performance 35*, 3 (2012), 424–438. 1

[Ste13] Steptoe W.: Camera requirements for a stereo rig for head-mounted video see-through ar for the oculus rift. modified logitech c310 cameras selected., 2013. 2

[UAW*03] Usoh M., Arthur K., Whitton M., Bastos R., Steed A., Brooks F., Slater M.: The visual cliff revisited: A virtual presence study on locomotion, 2003. 1, 4